Genetic Phylogenetic Analysis of PRRSV ORF5 Gene Based on a Geometrical Tool

Hua Tang, Xin-Sheng Liu, Yu-Zhen Fang, Li Pan, Zhong-Wang Zhang, Peng Zhou, Jian-Liang Lv, Shou-Tian Jiang, Wen-Fa Hu, Pan Zhang, Yong-Guang Zhang and Yong-Lu Wang
State Key Laboratory of Veterinary Etiological Biology, National Foot and Mouth Disease Reference Laboratory, Key Laboratory of Animal Virology of Ministry of Agriculture, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou 730046, China

Corresponding Author: Yong-Lu Wang and Yong-Guang Zhang, State Key Laboratory of Veterinary Etiological Biology, National Foot and Mouth Disease Reference Laboratory, Key Laboratory of Animal Virology of Ministry of Agriculture, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou 730046, China Tel/Fax: +86-931-8343796

ABSTRACT

The Z-curve is a geometrical tool for visualizing and comparing genomes. Since, the curve contains the information carried by the given sequence, DNA sequences could be analyzed systematically. In this study, the ORF5 gene of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) was analyzed by the Z-curve method and the DNAStar (DNASTAR Inc.) computer programs. The results revealed by these two methods were identical. The phylogenetic tree is more digitized and the Z-curve method is more picturesque. It turned out to be that, the Z-curve method shows a widely application prospect in phylogenetic relationships analysis. However, the Z-curve method is still in its premature stage. Novel algorithms are expected to be exploited to extract more information involved in the Z-curves.

Key words: Z-curve, geometrical study, PRRSV, ORF5, genetic diversity

INTRODUCTION

Analyzing DNA sequences and amino acid sequences with the aid of geometric approach is a novel way which shows the merits of intuitive, visual, vivid (Chu et al., 2009). Some people have applied geometric approaches into study, such as Li et al. (2010) and Giacomo et al. (2007).

According to Zhang and Zhang (2003) and Wen and Zhang (2003), Z-curve is a geometric tool to study the DNA sequence. Since, it was first proposed, the Z-curve has been extensively applied to life science research, such as, sequence segmentation (Zhang and Zhang, 2003; Wen and Zhang, 2003), horizontal gene transfer detection (Zhang and Zhang, 2003), isochoric domain inference (Zhang and Zhang, 2003; Wen and Zhang, 2003) and sequence analysis, for instance, nucleotide distribution analysis (Ou et al., 2003), replication origins Identification (Zhang and Wang, 2005), protein coding genes recognition (Zhang and Wang, 2000), separate base usages of genes revelation (Guo and Yu, 2007) and so on.

Porcine Reproductive and Respiratory Syndrome (PRRS) disease is caused by a positive single-stranded polyadenylated RNA virus (PRRSV) which belongs to Arteriviridae virus family in the order Nidovirales (Conzelmann et al., 1993). And its genome contains eight Open Reading
Frames (ORFs) which is about 15 kD (Cavanagh, 1997). Bred gilts and sows infected PRRSV may display abortion and premature farrowing and the nosogenic liveborn piglets may show respiratory tract disease (Cheon and Chae, 1999).

The PRRSV ORFs were consisted of 1a, 1b, 2B7 (Grebennikova et al., 2004). ORF1a and 1b are believed to encode nonstructural proteins, involved with viral replication and transcription (Grebennikova et al., 2004). The ORFs 2B7 encode the PRRSV structural proteins (Grebennikova et al., 2004). An effective way to distinguish wild type and vaccine-like strains of PRRSV is analyzing the Restriction Fragment Length Polymorphism (RFLP) of ORF 5 (Gagnon and Dea, 1998; Wesley et al., 1998; Cheon and Chae, 2000). Restriction sites diversities are awfully useful in diagnosis of disease (Cheon and Chae, 2004). GP5, the major viral envelope protein, encoded by ORF 5 is the most variable viral structural protein (Mardassi et al., 1995) and it is relevant to virus neutralization (Zhou et al., 2009). A variety of vaccines associating with GP5 were developed and performed well (Duran et al., 1997; Qiu et al., 2005; Xue et al., 2004).

Gaining insight into the information carried by the genomes of PRRSV would be helpful to discover drugs, develop vaccines as well as monitor their evolution and spread (Grebennikova et al., 2004). According to the genetic analysis of PRRSV isolates, PRRSV was divided into two predominant genotypes, the European and North American genotypes (Grebennikova et al., 2004; Lee et al., 2006). The ORF5 gene of PRRSV have been widely analyzed and researchers found it was useful to help us understand the genetic diversity of PRRSV isolates (Cha et al., 2006; Chen et al., 2006; Indik et al., 2000, 2005; Mateu et al., 2006; Stadejek et al., 2006; Thanawongnuwech et al., 2004; Zhou et al., 2009). In this study, the complete sequences of the ORF5 gene of the twenty-one PRRSV isolates were analyzed with the DNAstar (DNASTAR Inc.) computer programs and the Z-curve method.

MATERIALS AND METHODS

The PRRSV ORF5 genes and annotated information (Table1) were downloaded at the web site (http://www.ncbi.nlm.nih.gov/).

The Z-curve coordinates compute software (Zplotter) was downloaded at the web site (http://tubic.tju.edu.cn/zcurve/). Users are recommended to use Z curve plotter Java applet version which can be available at the web site (http://tubic.tju.edu.cn/zcurve/). However, for those who experience difficulty in using Java applet, or want to compute Z-curve coordinates locally, this local version of Zplotter can be used. The theory of Z-curve was reported by Zhang and Zhang (2003).

RESULTS

Twenty-one PRRSV isolates ORF5 gene were performed genetic phylogenetic analysis to reveal the genetic relationship and evolution of these PRRSV isolates with the help of DNAstar program (Fig. 1). Six distinct clusters were showed in the phylogenetic tree and we marked them A, B, C, D, E, F, respectively. And it is obviously that, there are two apparent genotypes among these twenty-one PRRSV isolates.

The twenty-one PRRSV isolates previously studied were then analyzed by the Z-curve method (Fig. 2). Figure 2a shows two evident genotypes of these virus the same as the result displayed in the phylogenetic tree. Figure 2b-d, respectively mapped the A cluster; B cluster and C, D, E, F
Table 1: The Zcurve plot marks, accession numbers, isolation regions of collected PRRSV and virus isolates cluster categories distinguished in the phylogenetic tree

<table>
<thead>
<tr>
<th>Z-curve plot mark</th>
<th>Accession number</th>
<th>Isolation region</th>
<th>Cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>e\005.1.txt' u 1:2:3</td>
<td>AB546121.1</td>
<td>Japan</td>
<td>A</td>
</tr>
<tr>
<td>e\005.2.txt' u 1:2:3</td>
<td>AF069503.1</td>
<td>Denmark</td>
<td>A</td>
</tr>
<tr>
<td>e\005.3.txt' u 1:2:3</td>
<td>AF069518.1</td>
<td>Denmark</td>
<td>A</td>
</tr>
<tr>
<td>e\005.4.txt' u 1:2:3</td>
<td>AF176426.1</td>
<td>USA</td>
<td>A</td>
</tr>
<tr>
<td>e\005.5.txt' u 1:2:3</td>
<td>AF176427.1</td>
<td>USA</td>
<td>A</td>
</tr>
<tr>
<td>e\005.6.txt' u 1:2:3</td>
<td>AF389496.1</td>
<td>USA</td>
<td>B</td>
</tr>
<tr>
<td>e\005.7.txt' u 1:2:3</td>
<td>AF389496.1</td>
<td>USA</td>
<td>B</td>
</tr>
<tr>
<td>e\005.8.txt' u 1:2:3</td>
<td>AF389496.1</td>
<td>USA</td>
<td>B</td>
</tr>
<tr>
<td>e\005.9.txt' u 1:2:3</td>
<td>AF389496.1</td>
<td>USA</td>
<td>B</td>
</tr>
<tr>
<td>e\005.10.txt' u 1:2:3</td>
<td>AY740010.1</td>
<td>Italy</td>
<td>C</td>
</tr>
<tr>
<td>e\005.11.txt' u 1:2:3</td>
<td>EU273703.1</td>
<td>Taiwan</td>
<td>D</td>
</tr>
<tr>
<td>e\005.12.txt' u 1:2:3</td>
<td>JF730974.1</td>
<td>Spain</td>
<td>C</td>
</tr>
<tr>
<td>e\005.13.txt' u 1:2:3</td>
<td>JN051728.1</td>
<td>Germany</td>
<td>C</td>
</tr>
<tr>
<td>e\005.14.txt' u 1:2:3</td>
<td>JN051741.1</td>
<td>Germany</td>
<td>B</td>
</tr>
<tr>
<td>e\005.15.txt' u 1:2:3</td>
<td>JN051746.1</td>
<td>Germany</td>
<td>A</td>
</tr>
<tr>
<td>e\005.16.txt' u 1:2:3</td>
<td>JQ917911.1</td>
<td>China</td>
<td>F</td>
</tr>
<tr>
<td>e\005.17.txt' u 1:2:3</td>
<td>U66579.1</td>
<td>USA</td>
<td>B</td>
</tr>
<tr>
<td>e\005.18.txt' u 1:2:3</td>
<td>U66581.1</td>
<td>USA</td>
<td>B</td>
</tr>
<tr>
<td>e\005.19.txt' u 1:2:3</td>
<td>U66587.1</td>
<td>Canada</td>
<td>A</td>
</tr>
<tr>
<td>e\005.20.txt' u 1:2:3</td>
<td>U66592.1</td>
<td>USA</td>
<td>E</td>
</tr>
<tr>
<td>e\005.21.txt' u 1:2:3</td>
<td>U66599.1</td>
<td>USA</td>
<td>A</td>
</tr>
</tbody>
</table>

Fig. 1: A phylogenetic tree based on the nucleotide sequence of the ORF5 gene of 21 PRRSV isolates. The tree was constructed with the aid of the DNASATR program. There are two apparent genotypes and six distinct clusters.
Fig. 2(a-d): The relevant plot marks for PRRSV isolates ORF5 gene used in this graph are listed in Table 1 (a) Portrays the three-dimensional Z-curves of all twenty-one PRRSV isolates ORF5 gene (AB546121.1, AF095503.1, AF176426.1, AF176427.1, AF339493.1, AF339494.1, AF339495.1, AF339496.1, AJ740010.1, EU273703.1, JF730974.1, JN651728.1, JN651741.1, JQ917911.1, U66379.1, U66381.1, U66387.1, U66392.1, U66399.1), (b-d), respectively portray the ORF5 gene of PRRSV isolates clusters (A; B; C, D, E, F) distinguished in the phylogenetic tree (Fig. 1), (b) Z-curves of AB546121.1, AF095503.1, AF095518.1, AF176426.1, AF176427.1, JN651746.1, U66387.1, U66399.1, (c) Z-curves of AF339493.1, AF339494.1, AF339495.1, AF339496.1, JN651741.1, U66379.1, U66381.1 (d) Z-curves of AJ740010.1, JF730974.1, JN651728.1, EU273703.1, U66392.1, JQ917911.1. That is to say, all the curves plotted in Fig. 2a are the union of curves graphed in Fig. 2-d. The legends of Fig. 2a are the union of legends plotted in Fig. 2b-d. As can be observed in Fig. 2a, the conspicuous red curve is the Z-curve of isolate JQ917911.1. It is obviously that, there are two apparent genotypes. From Fig. 2b-d, We can see that the differences between each Z-curve are remarkable. There is no doubt that the results revealed by Z-curve were identical with the sequences disclosed by the DNASTAR program.

cluster PRRSV isolates which were distinguished in the phylogenetic tree. There is no doubt that the results revealed by these two methods were identical. The phylogenetic tree is more digitized and the Z-curve method is more picturesque.
DISCUSSION

We can study the molecular phylogenies which may reveal many aspects of the transmission, epidemiology and evolution of rapidly evolving pathogens through genome analyzing (Pybus and Rambaut, 2009). Viral diversity, modulated by natural selection, global patterns of virus circulation and host population biology, is generated by evolutionary dynamics which resulted from a complex combination of rapid mutation and genome segment reassortment (Baillie et al., 2012). So the gene sequence study plays a key role in the life science.

It turned out to be that PRRSV could be gene typed through analyzing the ORF5 gene which was in agreement with Greiser-Wilke et al. (2010). Two genotypes of PRRSV were previously reported: American- and European-type strains (Indik et al., 2005; Nelsen et al., 1999; Zhou et al., 2011). And this was in accordance with the results of this study. And it was obviously that the ORF5 gene of PRRSV is highly variable (An et al., 2011). The DNASTAR program and the Z-curve method in this paper showed a identical accuracy in the genetic phylogenetic analysis. The former is more digitized and the Z-curve method is more pictorial. However, the phylogenetic tree could discriminate the genetic diversity more easily which can be proved in the pictures. If we want to confirm a virus belongs to which genotypes according to the Z-curve method we should accomplish a preparation work. The precondition is that we should construct a data base which definitely records the Z-curve coordinates area of different genotypes. Thus, if we plot a virus genome in a clearly marked genotypes Z-curve rectangular space coordinate systems, we could immediately read out what kind of genotypes the targeted virus is. So the Z-curve method is another reliably phylogenetic analysis tool. In addition, the Z-curve also can be used to recognize protein coding genes. Z-curve-E, a new self-training system derived from Z-curve, has been used to recognize protein coding genes (Guo and Zhang, 2006).

In general, Z-curve method shows its widely application in the areas of life sciences. For instance, gene sequence analysis, molecular epidemiology analysis, homology analysis, evolution analysis, phylogenetic relationships analysis, genetic diversity analysis. Currently, the Z-curve method is still in its premature stage. Since the more information is extracted from the Z-curve, the more accurate result can be gained, we are expected to explore more about the Z-curve. Gaining insight into the information carried by the genomes of viruses would be helpful to discover drugs, develop vaccines as well as monitor their evolution and spread. Some other novel algorithms are expected to be exploited to extract more information contained in the Z-curves.

ACKNOWLEDGMENTS

This study was financially supported by the National Pig Industrial System (CARS-36-06B) and Ministry of Agriculture’s Special Funds for Scientific Research on Public Causes, Research and Demonstration on evaluation technologies of clinical immune responses of serious animal diseases vaccine, No. 201203039).

REFERENCES


